

基于数据驱动的群智感知任务分配算法

张震, 李鹏

(国家计算机网络应急技术处理协调中心, 北京 100029)

摘要: 群智感知技术的应用实现了人群感测作用的最大化, 作为社会网络研究的核心技术之一, 然而对于感知参与者的位置和轨迹不确定性的问题造成群智感知数据实时性较差。为此, 提出了一种基于空间任务分配的移动群智任务分配算法, 该算法采用动态和自适应的数据驱动方案获取最优的模式来解决感知动态化问题; 算法基于公开历史轨迹的移动模型(基于马尔可夫模型), 根据初始任务按照贝叶斯推理来估算下一位置, 基于该算法的数据采集策略可以实现有本地服务引导未来数据的收集, 从而完成整个感知的回路反馈。所提出的任务分配被证明基于不确定轨迹的移动群智感知任务分配是有效的。

关键词: 群智感知; 社会网络; 自适应的数据驱动; 任务分配

中图分类号: TP301 **文献标志码:** A **文章编号:** 1001-3695(2017)08-2376-04

doi: 10.3969/j.issn.1001-3695.2017.08.032

Crowd sensing task assignment algorithm based on dynamic data driven

Zhang Zhen, Li Peng

(National Computer Network Emergency Response Technical Team/Coordination Center of China, Beijing 100029, China)

Abstract: The crowd sensing has realizing the maximization perception for crowd, which has been considered the key technology for society network. However, to perceive the position of the participants and trajectory uncertainty problem caused by poor group of mental perception data real-time. This paper proposed a crowd sensing group of intellectual task allocation algorithm based on space task allocation, the algorithm adopted the dynamic and adaptive data-driven scheme to obtain the optimal model to solve the problem of dynamic perception; Track algorithm based on public mobile model (based on Markov model), Bayesian inference, according to the initial task to estimate the next position, data acquisition strategy based on this algorithm could achieve a local service to guide the future data collection, thus complete the entire loop feedback of perception. Task assignment is proved based on the uncertain path of mobile group of mental perception task allocation is effective.

Key words: crowd sensing; society network; adaptive data driven; task assignment

0 引言

基于移动互联网的群智感知技术已成为学术界研究的热点课题, 所谓群智感知是指独立个体通过智能传感设备(智能手机或其他智能设备)共享一些零散资源(如图片、视频、音频、位置信息、地理空间信息等), 服务商通过智能分配算法对这些资源进行整合形成最终的处理结果, 当前学术界将智能分配算法的研究作为群智感知系统优劣的评价标准之一^[1,2], 对于智能分配算法的研究需遵循以下三个原则: a) 感知数据的来源需遵循个体独立性原则; b) 系统终端用户所被分配的数据均来源于 a) 中的感知数据, 并且遵循原则 a); c) 系统所分配的数据均用来服务最终的应用, 并且在特定的体系中, 终端用户和独立个体可互换角色。

当前对于群智感知的应用主要集中在地理空间信息感知方面, 该应用主要用于对一些恐怖事件、地质灾害以及特殊的一些行为进行预测。例如在事件预测应用的实例中, 系统可以根据参与者所感知周围的地理空间信息数据进行分析预测, 系统根据相应的任务分配算法进行预测过程的分析, 对于任务分配算法的研究主要有基于事件规模^[3]、基于预测开销^[4]等作为任务分配所遵循的原则。这些应用研究中, 由于参与感知人

员的移动性, 造成任务分配计算的结果的实时性较差, 如图 1 所示为空间任务分配算法应用的实例介绍。

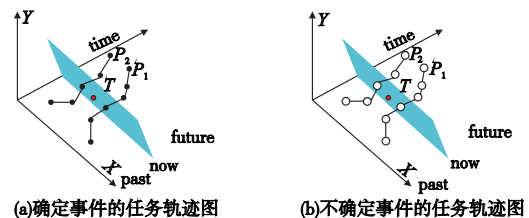


图1 群智感知例图

如图 1 所示, P_1 和 P_2 分别为任务分配算法在不同时刻分配的事件的任务曲线图, 其中时刻 T 表示为当前发生的事件点, (a) 为确定事件的任务轨迹图, 而 (b) 为基于不确定事件的任务轨迹图。假设基于图 1 中的两类任务轨迹进行算法运算, 图 1(a) 中参与者能获得一个确定的任务结果, 而图 (b) 中由于轨迹的不确定性造成任务分配结果的最终不确定性。因此设计实现基于动态任务分配的群智感知算法, 在提高感知范围、降低感知成本的前提下用于解决图 (b) 中所面临的动态、地理位置信息不确定的问题。

本文针对在移动群智感知的应用研究中, 面临非确定性的地理位置信息的任务分配问题, 基于 Darema 等人^[5]研究的

收稿日期: 2016-06-03; 修回日期: 2016-08-01

作者简介: 张震(1984-), 男, 博士, 主要研究方向为语音信号处理、语音识别、关键词检索(kingksl@126.com)。

DDAS系统的部分思想进行演进,该系统的核心思想是将群智感知应用中所面临的较大数据分配问题进行数据重组、分步计算。然而这些任务分配的应用研究均存在如下问题:a)感知数据提供者所提供并非当前时刻的数据,造成实际群智感知应用出现预测偏差的问题;b)应用系统最终的输入数据是通过相应模型对所收集的数据处理后的重组数据,缺乏数据的可靠性;c)感知参与者的位置不确定性问题,缺乏移动模型的支持。本文设计实现了基于动态数据采集、融合以及分析,并形成反馈回路的群智感知任务分配算法。

1 相关研究工作

1.1 群智感知任务分配

群智感知的思想早在互联网兴起时就被提出,如SETI项目就是利用互联网思维聚合全球各地的计算资源实现一个大规模计算的任务,这是最早的一种基于互联网的任务分配应用。当前随着移动互联网和云计算的快速发展,在群智感知研究中,任务分配是研究的核心点之一,当前国内外学者对于任务分配的研究主要分为任务自主选择和任务协调分配两类。其中在任务自主选择的方案研究中,文献[6,7]从任务池中动态地给各个感知参与者自动分配相应的任务(基于相关的自组分配算法),该策略由于所选的任务并非全局优化后的任务,所以在传感成本和全局任务分配上存在效率低下的问题,而本文提出的任务分配解决策略是采用马尔可夫概率链的方式,为感知参与者提供最优的、效率最高的任务;在任务协调分配的研究中,通常通过对合格的感知参与者分配可同的传感资源直至应用目标的满足,然而应用目标往往存在标准不一致的情况,因此对任务分配的协调算法也有较高的要求^[8-10],而本文提出的感知用户分配中依然采用概率论的模式基础,为各个感知参与用户提供最优的任务分配。本文的研究重点基于任务协调分配,侧重对参与者的不确定轨迹和缺乏确定位置的结果预测(应用DDAS思想解释动态数据及应用需求)。

1.2 移动建模和预测

近来,对于移动性研究主要针对采用高分辨率定位手段实现个人移动模式和构建流动模型的数据挖掘模型的设计^[11,12]。相比基于随机生成运动的合成流模型相比,基于跟踪迁移模型更加适合对于个人统计模型对真实世界反映的仿真,这类模型通常基于步行移动的追踪,极少考虑其他的移动模式(如驾车或者其他交通工具)以较低模型因素的复杂性。通常基于位置预测的解决采用马尔可夫模型^[13],或者采用基于历史轨迹的贝叶斯推理预测解决方案^[14]。本文引入以上流动位置预测模型为应用得到任务分配参与者更为准确的位置信息。

基于以数据为中心的群智感知任务分配的研究,是以移动模型为算法基础,按照协同分配、智能规划为各个参与者分配最优的任务配置。

马华东等人^[15]提出了一种应用于移动网络的机会路由策略,介绍了移动机会网络面临的挑战,详细阐述了机会路由算法的评价指标、设计需求与转发机制,从信息扩散机理、能量有效的数据转发策略、节点社会性辅助的路由机制等三个方面介绍了相关的研究成果,基于移动互联网的机会路由算法对群智感知任务分配提供了一个可靠的理论传输基础。

2 任务分配的动态数据驱动框架设计

本文基于DDAS框架提出一种新的基于不确定归集的动态任务分配策略,如图2所示为本文提出的任务分配策略的框架图,框架中包含参与部分、分配服务以及应用三部分。其中参与部分完成感知参与者对于应用所需数据的收集以及位置的确定;分配服务通过移动数据模型(泛洪移动模型)进行数据的处理以及任务分配的计算处理,其中处理过程是通过移动模型动态地过滤数据,实现数据的增强或补充,用于预测或模拟基于历史轨迹的参与者的当前位置;应用在对更新后的目标和需求以及参与者的信息和预估的位置,将该信息作为任务分配模块的输入参数,按照概率模型将最终的任务分配给相应使用者;当感知参与者获取相应任务后,比如应用需获取其确切位置;最后的任务分配用来引导未来数据的收集,完成应用所需前往目标定位确切位置。该框架以离线学习过程的移动模型作为基础,通过基于移动模型的流程模块按照公共轨迹进行大数据处理。

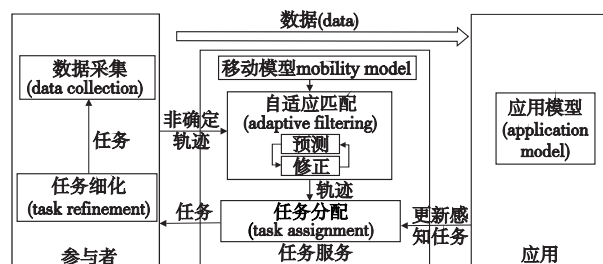


图2 基于空间任务分配的自适应动态数据驱动框架结构图

由于应用程序所需处理的模型依赖于相关特定的应用程序和感知任务,以下将对框架中的各个组件进行详细介绍。

2.1 自学习移动模型

自学习移动模型,本文通过采用公开的历史轨迹数据作为模型输入参数,通过该模型计算相邻位置转移的概率,在概率模型的选择中,采用基于贝叶斯推理和马尔可夫两类经典概率论预测模型。在模型设计过程中不考虑参与者所搭乘不同交通模式的因素,然而自学习移动模型对于该因素的扩展是较简单的,本文将不作详细介绍。

在研究移动模型中,历史轨迹的采集至关重要,这也是预测学研究的重点,根据历史轨迹来预测未来的个人可能发生的轨迹点(该方面的研究需要确保个人隐私的保护),以下对两类经典的概率预测模型应用到本模型的详细分析。

2.1.1 贝叶斯推理

贝叶斯推理的架构可以通过一个网格图来形容道路的网络轨迹映射的预测。通常在网格图研究中,将地图作为一个 $G \times G$ 粒度的二维网格(所有的位置点作为相同的粒子点),基于不同网格单元构成一个整体地图,每个单元可作为图中的一个节点,用节点序列来表示相关轨迹的映射。

通过相应轨迹序列映射可形成相应的轨迹有向图,图3(b)中 T_1 、 T_2 等轨迹同3(a)中一致,通过点来表示图3(a)中对应的网格图象限。通过轨迹化后的模型结构可用基于概率的模型结构进行描述,如节点 n 的下一个目的地为轨迹点 T_j 。

定义式(1)表示节点 n 根据历史轨迹预测下一轨迹的概率。

$$P(\text{next}(T_j) = n | T_j) = \frac{P(T_j | \text{next}(T_j) = n) P(\text{next}(T_j) = n)}{P(T_j)} \quad (1)$$

同样根据模型中总节点数 n 来预测下一轨迹如式(2)所示。

$$P(\text{next}(T_j) = n) = \frac{|\{T_i | (T_j, n) \in T_i\}|}{|T|} \quad (2)$$

2.1.2 马尔可夫模型

采用如图3(b)中相同的模型结构,构造马尔可夫模型如图3(c)所示,图中表示每个节点的状态图,对于每一个节点的转换条件 P_{rs} 和 P_{sr} 分别表示节点 n_r 和 n_s 的相互条件转换概率,如式(3)所示为节点序列 (n_r, n_s) 的一阶马尔可夫转换概率。

$$P_{rs} = \frac{|\{T_i | (n_r, n_s) \in T_i\}|}{|\{T_i | n_r \in T_i\}|} \quad (3)$$

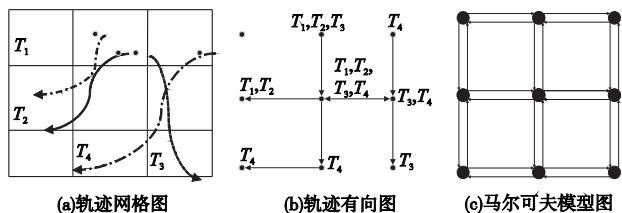


图3 网格、有向图、马尔可夫模型表示轨迹展现图

通过以上贝叶斯推理和马尔可夫模型对预测概率的分析,本文所构建的预测模型完全基于马尔可夫概率论原理,接受多变量输入,并且与动态数据任务分配的框架能够柔性集成。

2.2 自适应组件

本文提出的动态任务分配框架中的自适应组件的作用是提供时间服务机制、位置精确性服务机制等,如图4所示模型中 X 状态表示群智感知实际运行状态节点, X_{i+1} 表示 X_i 按照流程执行的下一个状态节点, Z 表示对状态 X 数据采集的信息集合(观察状态),定义式(4)表示图4中的模型。

$$x_{i+1} = Ax_i + \omega_i \quad \omega_i \sim N(0, Q) \quad (4)$$

其中: A 为常数,表示线性模型的噪声值,其服从正太分布 $(0, Q)$,根据马尔可夫模型定义的前一个状态为,将实际采集过程中存在的高斯白噪声 ω ,根据过程噪声方差 q 定义各个状态 x_i 相应的观察状态 z_i ,定义观察模型方程为。

$$Z_i = Hx_i + v_i \quad v_i \sim N(0, R) \quad (5)$$

其中: H 为线性测量噪声系数,表示噪声修正值,其服从正太分布。在实际系统的应用中,按照不同的建模需求,对于噪声的设置采取设备自身参数(如GPS噪声)或者微扰法(如差分法)等方式。本文假设存在一个高斯噪声方差 R ,为了获取真实状态的信息,需要设置相应的滤波算法对噪声信息进行处理,本文采用卡尔曼滤波器和粒子滤波器作为滤波处理的手段(分别针对不同的噪声,如卡尔曼滤波器可对线性过程的高斯噪声产生最优过滤效果,而粒子滤波过程可对状态空间模型的假设模型或噪声进行滤波处理)。

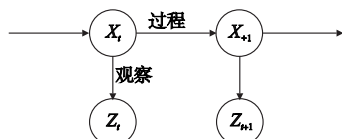


图4 简单的状态-空间模型

2.3 非确定性的空间任务分配策略

构建如下模型:定义在时间 T 存在一组位置不确定的空间任务分配需求,应用程序可更新空间任务和目标任务,参与者对应相关的不确定性位置信息。根据模型结构算法可对该不确定性以及作业优化过程进行处理,并获取所需目标信息。

任务分配策略描述如下:

a)任务分配可映射为参与者的一个任务,对于相关任务-参与者的匹配可划分为一个单独的任务模块,并且每一个任务均需要定义其成本值;

b)本文定义的空间任务包含一个目标(一个物体、事件或者现象)、一个位置、一个基于时间序列的传感表、自定义的消息指令(包含数据传输过程中涉及到的控制指令);

c)本文对于感知参与者参与感知的评估指标采用距离的成本预算模型,采用业界认可的距离成本定义标准,并规定成本最低的作为此次任务的参与者;

d)任务的范围采取如下不同的方式定义:

(a)单范围模型,指每次任务只需要一个参与者参与执行即可完成;

(b) K 范围模型,指每次任务需要 K 个参与者完成,此模型用于非理想环境(存在数据丢包的情形)。

基于位置不确定的任务分配策略:对于实际环境下的群智感知数据采集,感知参与者的位置不确定性是往往会造成应用服务的不可用,如何对位置的处理是群智感知服务研究的重点。当前,学术界将该问题的研究称为时空查询的不确定性问题,并出现多项研究成果,如 Nearest Neighbor、Top-K 以及 Range Queries^[16]等算法,这些基于查询的算法主要针对的对象仅为单个分配结果,因此无法直接应用于本文所提出的分配策略中,本文的任务分配查询策略描述如下(任务的获取)。

1. Set the uncertainty locations of participants as a set Q , and using a probability function $f(x)$ to describe Q ;
2. For each participant:
 - we calculate the centroid point of all possible location samples in each uncertainty area and use it to calculate the expected distances between this point P and all locations of the task set T ;
 - OR
 - we first apply a geometric pruning algorithm to remove the task-participant pairs with zero probability of being accessible (i. e. a participant can't travel to the place of a task) and shrink the uncertain areas to only contain the accessible location samples;
3. For the remaining pairs with shrunk areas:
 - we calculate the probabilities of the task locations being accessible by participants as well as the expected distance between task locations and shrunk areas;
4. The set of expected distances and the accessibility probabilities can be used in our probabilistic task assignment methods proposed in paper^[17].

任务的本地分配:通过以上对于不确定位置的任务获取描述,任务分配服务端即完成具体的任务分配过程,并分发到个体参与者。本地任务细分的目的是为了进一步优化任务分配(确保参与者能够获取服务端的确定位置,通过相对矢量运算可获取自身的确切位置,这样服务端就能获取每个参与者的实时确切位置,这样的细化是以增加硬件开销为代价的)。最后服务端根据运算结果将具体的任务分配给相关参与者完成未来数据的引导收集过程,从而完成了整个群智感知的回路反馈。

本文所提出的基于动态数据驱动的群智感知任务分配策略基于以上整个反馈循环过程,采取针对性的数据收集手段,动态地适应应用需求以及数据的获取,并且根据参与者的移动轨迹和感知任务开销成本作为分配参考标准,使得资源利用率得到了更大的优化。

2.4 实验验证

本章节通过实验室场景对本文所提出的基于数据驱动的

群智感知任务分配算法,实验通过基于特殊环境下的参与者(不考虑网络传输的影响)对服务分配的任务信息反馈的准确率来衡量算法的有效性。

实验环境描述如下,实验参与者为100个体,服务端运行文献[16]中的Top-K算法以及本文所提出的动态数据驱动的任务分配算法(分时段调用),实验结果显示,在实验周期(16 h)内通过更迭算法交换的方式对感知数据与真实数据作比较。

如图5所示为实验结果统计分析图,在6~10时刻内为执行本文提出的动态数据驱动的任务分配算法,通过与真实数据对比,感知数据的准确率在95%以上,相比Top-K算法的(其余时刻调度为该算法)50%左右准确率有显著的提升。

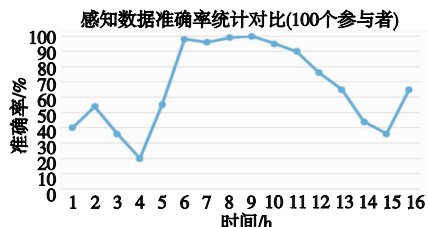


图5 感知数据准确率统计对比图

算法的复杂度验证通过服务端资源消耗率的使用进行统计,如图6所示,本算法的应用相比Top-K算法明显增加了资源的消耗量,这是本算法后期改进的重点之一。

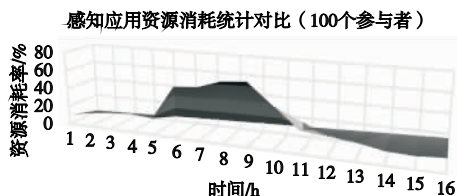


图6 感知应用资源消耗统计对比图

3 结束语

群智感知技术的发展是建立在参与者精确地感知数据基础上的,这就要求服务端为个人移动平台分配有效的空间任务。然而由于感知参与者所处的动态环境与位置不确定性,学术界提出了用于解决动态空间任务分配框架的DDDAS模式。本文针对这种环境提出了如下策略:a)建立动态数据驱动框架与反馈回路,用于数据采集以及实时适应动力学的数据,以便适应应用程序的需求;b)利用自适应滤波的机制原理,通过参与者的公共轨迹来不断地调整参与者动态以及不确定位置的精确定位;c)任务分配算法选择最优的(采用最大覆盖或者最低成本)一组参与者实现特定于应用程序的目标任务。

下期要目

❖ 虚拟手术中软组织切割模型研究进展

❖ 机械臂绝对定位精度标定关键技术综述

❖ 信任驱动的云联盟博弈算法

❖ 基于社团结构的节点的影响力分析

❖ 点目标和扩展目标联合跟踪算法

❖ 嵌入式系统源程序级软件能耗建模与分析

❖ 网络视频服务提供模式与仿真研究

❖ 基于网络模型的城市公共自行车需求量预测研究

❖ 一种改进EM算法的跨领域情感分类方法

参考文献:

- [1] 赵东. 移动群智感知网络中数据收集与激励机制研究[D]. 北京: 北京邮电大学, 2014.
- [2] 杜扬, 黄河, 孙玉娥, 等. 地理位置相关移动感知系统任务分配问题研究[J]. 计算机研究与发展, 2014, 51(11): 2374-2381.
- [3] Kazemi L, Shahabi C. Geocrowd: enabling query answering with spatial crowdsourcing[C]//Proc of the 20th International Conference on Advances in Geographic Information Systems. 2012: 189-198.
- [4] Pournajaf L, Li Xiong, Sunderam V, et al. Spatial task assignment for crowd sensing with cloaked locations[C]//Proc of International Conference on Mobile Data Management. 2014.
- [5] Darema F. Dynamic data driven applications systems: a new paradigm for application simulations and measurements[C]//Proc of Computational Science. [S. l.]: Springer, 2004: 662-669.
- [6] Shin M, Cornelius C, Peebles D, et al. Anonymsense: a system for anonymous opportunistic sensing[J]. Journal of Pervasive and Mobile Computing, 2010, 7(1): 16-30.
- [7] Deng Dingxiong, Shahabi C, Demiryurek U. Maximizing the number of worker's self-selected tasks in spatial crowdsourcing[C]//Proc of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2013: 314-323.
- [8] Das T, Mohan P, Padmanabhan V N, et al. Platform for remote sensing using smartphones[C]//Proc of the 8th ACM International Conference on Mobile Systems, Applications, and Services. 2010: 63-76.
- [9] Reddy S, Estrin D, Srivastava M. Recruitment framework for participatory sensing data collections[C]//Proc of the 8th International Conference on Pervasive Computing. Berlin: Springer, 2010: 138-155.
- [10] Riahi M, Papaioannou T G, Trummer I, et al. Utility-driven data acquisition in participatory sensing[C]//Proc of the 16th International Conference on Extending Database Technology. 2013: 151-162.
- [11] Lin Miao, Hsu Wenjing. Mining GPS data for mobility patterns: a survey[J]. Pervasive and Mobile Computing, 2013, 12(6): 1-16.
- [12] Musolesi M, Mascolo C. Mobility models for systems evaluation[M]//Middleware for Network Eccentric and Mobile Applications. [S. l.]: Springer, 2009: 43-62.
- [13] Liao Lin, Patterson D J, Fox D, et al. Learning and inferring transportation routines[J]. Artificial Intelligence, 2007, 171(5): 311-331.
- [14] Xue A Y, Zhang Rui, Zheng Yu, et al. Destination prediction by sub-trajectory synthesis and privacy protection against such prediction[C]//Proc of the 29th International Conference on Data Engineering. 2013.
- [15] 马华东, 袁培燕, 赵东. 移动机会网络路由问题研究进展[J]. 软件学报, 2015, 26(3): 600-616.
- [16] Wang Yijie, Li Xiaoyong, Li Xiaoling, et al. A survey of queries over uncertain data[J]. Knowledge and Information Systems, 2013, 37(3): 1-46.
- [17] Pournajaf L, Li Xiong, Vaidy S, et al. Spatial task assignment for crowd sensing with cloaked locations[C]//Proc of International Conference on Mobile Data Management. 2014.
- [18] 黄涵霞, 丁强, 李莉, 等. 移动终端智能引擎平台研究[J]. 计算机研究与发展, 2014, 24(6): 184-186.

❖ 一种基于质量估算的空间数据流聚类算法研究

❖ 面向VANET可靠广播的协同背包算法研究

❖ 基于能效的OFDMA系统资源分配研究

❖ 基于ISAP的D2D资源分配优化算法研究

❖ 两类具有极低自相关的二元序列

❖ 基于GHM多小波算法的功耗分析攻击

❖ 基于DTW算法的旁路功耗信号动态伸缩对齐

❖ 基于压缩域时空复杂度的实时视频满意度模型

❖ 视频图像的车辆速度实时检测算法研究

❖ 基于总变分的MPEG解码算法